Specialized lexicography by means of a conceptual data base: establishing the format for a multilingual marine dictionary

Jean-Luc Husson, Nadia Viscogliosi, Laurent Romary, Sylviane Descotte, Marc van Campenhoudt

Specialized lexicography by means of a conceptual data base: establishing the format for a multilingual marine dictionary

Second Conference on Maritime Terminology
University of Turku, Finland
11-12 May 2000

Jean-Luc Husson <Jean-Luc.Husson@loria.fr>
Nadia Viscogliosi <Nadia.Viscogliosi@loria.fr>
Laurent Romary <Laurent.Romary@loria.fr>

Loria (UMR 7503)
Équipe Langue et Dialogue
BP239
F-54506 Vandoeuvre-lès-Nancy,
FRANCE

Sylviane Descotte <sdescotte@isti.be>
Marc Van Campenhoudt
<marc.van.campenhoudt@euronet.be>

Centre de recherche Termisti
Institut supérieur de traducteurs et interprètes
34, rue J. Hazard
B-1180 Brussels
Belgium

Abstract

The *Hydrographic Dictionary* published by the International Hydrographic Bureau (Monaco) was computerized as part of the European research project *Dhydro*. The aim of the project was to develop a multilingual electronic dictionary with an interactive editorial infrastructure based on the new information technologies.

This paper will firstly provide a summary of the analysis that has resulted in the hitherto followed lexicographical approach being abadoned in favour of a monosemic perspective more suited for the management of a large number of source and target languages. This will be followed by a detailed description of the strategy that has enabled the tagging of the surface structure of each monolingual volume of the dictionary in accordance with the prescriptions of the *Text Encoding Initiative,* and the subsequent merging of all data in one single multilingual data base which conforms to the ISO 12 200 standard on the exchange of terminological data.

Finally, the authors will show how XML records resulting from the retroconversion process will enable the reproduction of numerous variants of the dictionary thanks to the use of stylesheets (*XSL Transformations*).

Key words: specialized lexicography, multilingual terminological data base, conceptual model, hydrography, XML, TEI, MARTIF.

1 Introduction

A preliminary study of the computerization of the *Hydrographic Dictionary* of the International Hydrographic Organization (IHO, Monaco) was presented at the *First International Conference of Maritime Terminology* (Bessero *et al*. 1999). Since then, the underlying objective of that paper has been pursued within the framework of the *Dhydro*[1] project which is financed by the European *Multilingual Information Society* (MLIS) programme.

While the initial analysis was essentially intended to save the data in the shape of a structured document, the various partners[2] within the *Dhydro* project had a more ambitious goal, i.e. to develop a multilingual electronic dictionary whose editorial infrastructure would be based on the new information technologies[3].

2 Difficulties related to the computerization of the *Hydrographic Dictionary*

At the outset, the *Hydrographic Dictionary* was a specialized lexicographic tool designed for English. Each new edition was adapted to other languages in the form of independent monolingual volumes in which terms are arranged alphabetically, with cross-references to the English entries by means of numbers (or pointers). The corpus of the latest English edition comprises 6,064 entries covering a large number of concepts that cannot be found in other marine dictionaries.

Bessero *et al.* (1999: 192-193) already showed that the system complicates the electronic management since any changes in the English version require the renumbering of the entries and the updating of the equivalences in the translated volumes. Although the linguistic analyses conducted as part of the *Dhydro* project underscored the wealth of the dictionary, they also showed that the referencing system of the printed versions was at once too rudimentary and too complex to enable a smooth resolution of the typical problems related to polysemy and the search for equivalent terms. The cases below give some idea of the difficulties encountered in the course of the computerization task.

2.1 A lack of international consultation for the definition of the underlying realities of the field

- The terms that are not described in English do not get a pointer to this language. For instance, the underlying meaning denoted by the French term *concordance* is not dealt with in the English volume [example 1][4].
- The same is true for acceptations that are not described in English. For instance, meaning b) of the French term *constellation* does not have a pointer to the English term *constellation* [example 2].
- A term from a language may refer to a closely related concept without the approximative nature of the equivalent being expressed. So, the French term *crachin* is attached to the English entry *drizzle* whose definition is much closer to that of the French *bruine* [example 3].

---

1. Abbreviation of *Dictionnaire hydrographique*, and a reference to the French encyclopaedist Diderot.

2. The International Hydrographic Bureau (IHB, Monaco), the Laboratoire lorrain de recherche en informatique et ses applications (LORIA, Nancy), the Service hydrographique et océanographique de la marine (Shom, Paris), the Centre de recherche TERMISTI (Institut supérieur de traducteurs et interprètes, Brussels), and the Institut für deutsche Sprache (IDS, Mannheim).

3. This second aspect has already been dealt with in another paper at the Conference.

4. The examples can be found in the annex.

- When there is no real translation available for a term, the other languages propose a translation by means of the equivalent of the encompassing term. For instance, the French *estran*, which is equivalent to the English *strand* (5074) also serves as a translation for the term *foreshore* (1907), which denotes a much more restriced semantic field [example 4].

2.2 Asymmetry and complexity of the referencing system

- The reference mechanism is not symmetrical since a precise and monosemic acceptation in the foreign language points to a polysemous English entry. Conversely, when polysemous entries appear in English, the individual meanings are not numbered, which complicates any attempt at translating the English into another language and requires the repeated use of the other volume [example 5].
- In the English volume, a phrase may refer to a main entry where it is broken down (e.g. *hydrographic signal, longitude signal,* and *luminous signal*), while the opposite also occurs (e.g. *eccentric signal, fog signal*, or *sound signal*) [example 6].
- In each volume, the synonyms are scattered within the alphabetical order of the entries and refer to the definition of a head term. The reference is logically to the English equivalent, but there are numerous complications. For instance, the French term *marée de tempête* in English refers both to the main term *storm surge* (5066), and its synonym *storm tide* (5067). In French, the same term refers to *onde de tempête*, which leads to *wave: storm* (5918), which, in turn, refers to *storm surge* (5066) [example 7].
- Synonyms may appear within a definition in a given language and as specific meanings within another. In example [8], the French volume includes the acceptation 'Synonyme de HAUTEUR a) ou b)" in the entry of *élévation* without, however, mentioning the equivalence – by extension through *hauteur* b) – with the English *altitude* (101).
- As a general comment, beyond the cases of synonymy, it must be said that the consultation and interpretation are highly complicated as certain concepts are described in other entries. In this respect one may cite the example of the French term *heure Zoulou*, which in French refers to meaning c) of *heure* – which does not have an equivalent – while there is also a reference to the English equivalent *zulu time* (6064), which, itself, refers to *time: Greenwich mean* (5522) [example 9].
- Analogous references of the type 'see' or 'see also' link the entries in English, whereas they are more accurate in French, where they link the meanings. This phenomenon is perhaps best illustrated by the mention 'Voir HEURE c)' in example [9].

The complexity of the few cases described above would logically prompt the discerning reader to tread cautiously and keep the English volume close to hand whenever the French or Spanish volumes[5] are consulted to translate between these two languages.

The most obvious solution when contemplating the computerization of a bilingual polysemous dictionary would seem to be to deal with equivalence through meaning-to-meaning pointers. This solution poses numerous problems in view of the increasing number of languages, with English perhaps no longer playing its pivotal role. The move towards an Internet-based interactive computerized management of the data base thus required an in-depth study as to which data model should be adopted. One of the main objectives of the new project was to enable the editors responsible for each of the languages to exchange points of view and negotiate definitions (i.e. a

5. In this paper no examples are given of the Spanish volume since this is a literal translation of the English version. The extracts from the French version pose typical problems related to the search for equivalents and are, as such, better examples.

division of meanings) continuously in order to ensure an improved equivalence of the terms. This also implied that the cutting up of the underlying reality and the definitions ensuing from it could be continuously revised without much difficulty.

3 From the book to the data model

Nancy Ide et Jean Véronis (1996: 174) rightly noted that dictionaries are at once texts and data bases, and «thus reveal a major dichotomy made up of their surface structure (the text) and their deep structure (the informational content).»

As a result of the initial analysis (Bessero *et al.* 1999: 183-186) it was decided to computerize data in conformity with the norms of the *Text Encoding Initiative (TEI)*[6], which are generally complied with in the computerization of texts in the Arts and Humanities. More specifically, the recommendation was to follow chapter 12, which is devoted to printed dictionaries, and, consequently, to adopt an editorial policy that was largely based on the preservation of the surface structure of the text (order of the entries and fields, numbering of the definitions, etc.).

As the *Dhydro* project aimed to design a multilingual dictionary in which each language can indiscriminately be used as target or source language, it became necessary to adopt a more 'lexical' approach. Since this is entirely independent of the layout of the dictionary, it would allow a description of the deep structure in the form of a perfectly structured data base. Within this new approach, the so-called 'conceptual' terminology management model seemed by far the most suitable since it had the advantage of retaining the successive description levels, each of which refers to the same monosemic meaning, called 'concept' for practical rather than philosophical reasons.

1 concept
    described in n languages
        denoted by n terms in each language

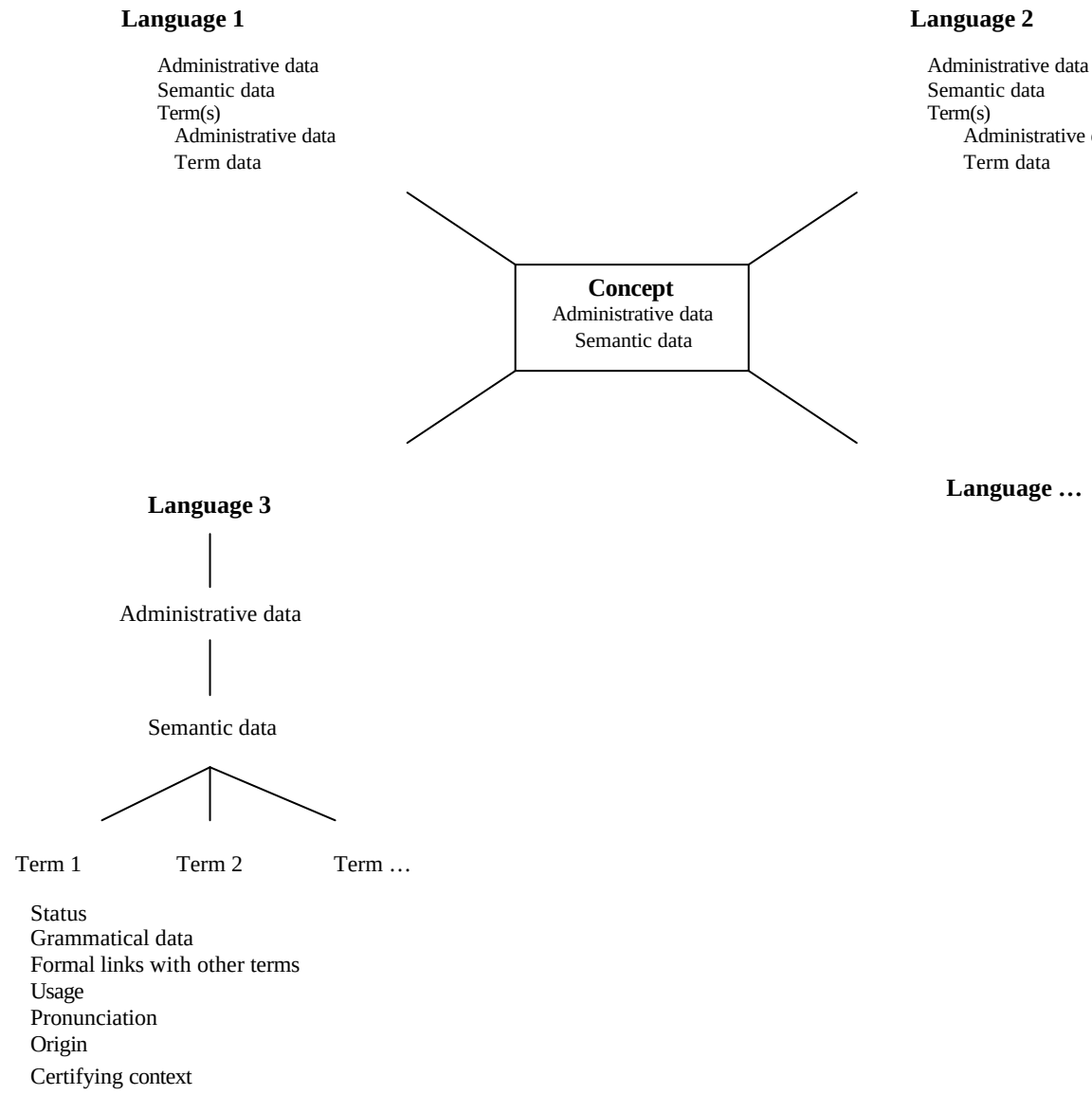
An in-depth linguistic study of the dictionary indeed leads to the conclusion that the wealth of information contained in it could be far better exploited if each entry were accurately identified by a distinct data field placed at the most appropriate level. Since a detailed overview of the content of each field falls outside the scope of this paper, the data model adopted is represented in the following synthetic diagram.[7]

While the prescriptions contained in chapter 12 of the TEI – already adopted by Bessero *et al.* 1999 – were a good starting point to tag the contents of each monolingual dictionary, the subsequent merger of the various volumes required compliance with the ISO 12 200 standard (1999). Better known as *MARTIF*, this norm constitutes a *de facto* development of chapter 13 of the TEI which is devoted to terminological data bases[8]. All the fields set forth for the new *Hydrographic Dictionary* perfectly correspond to those contained in this standard, as well as in ISO 12 620, on which it is based.

6. http://www.tei-c.org/

7. Details on the complete model can be found in three languages on the *Dhydro* web site: http://www.loria.fr/projets/MLIS/DHYDRO/

8. There is a competing format, known as Geneter. This has not been used since from the start the *Dhydro* project was to some extent an extension of the activities of the TEI.



4 The retroconversion of the printed versions

One of the main tasks within the project consisted of merging the English, Spanish, and French volumes into a single multilingual dictionary. Thanks to a very rigorous methodology this delicate task has been accomplished in two stages. The first involved the SGML tagging of the monolingual volumes in conformity with the TEI. This was followed by XML[9] tagging of the multilingual dictionary in accordance with the MARTIF standard.

9. A recent development in the world of tagging languages is the emergence of XML as a standard, with its recommended use for the creation and distribution of structured documents on the WWW. XML is a simplication of the SGML norm and allows easier use of structured documents within a client-server type application. More details and background on the choice is given in our second contribution to this Conference.

4.1 From text format to TEI encoding

The three base files (*Word, Wordperfect*) were first converted into HTML files in order to retain the tagging of the typographic features (tabs, spaces, capitals, bold or italic characters, etc.).

The bulk of the recurrent information to be encoded was retrieved either through linguistic markers (e.g. 'see also', 'the opposite is', 'on dit aussi', 'terme obsolète', etc.), or by the typographical style linked to them in accordance with automatic search mechanisms. This resulted in the marking of synonyms, antonyms, associated terms, and indications of usage.

However, the base files required a great deal of research and manual corrections (errors, lacunae, file anomalies, lack of homogeneity). The sometimes random aspect of this method explains why certain elements could not be encoded as they were formulated in an uncommon manner. For instance, the definition of the French term *goulet* actually contains two meanings, as well as an indication of the frequency of said term in the second meaning. Though untagged in this stage of the encoding process, the information may nevertheless be extracted and represented later in the final data format.

```
<entry id="fr-goulet">
<form type="main"><orth type="main">goulet</orth></form>
<sense n="1" id="fr2072s1">
<def>Passage étroit faisant communiquer un port ou une rade fermée avec la mer. Ou plus rarement passage navigable étroit d'une baie, d'un détroit, d'un fleuve, etc.</def>
</sense>
</entry>
```

After the encoding stage, entries were numbered, which, at first, involved a combination of the language and the term. The numbering was an important parameter which had to be taken into account as early as possible since this identifier allowed the unambiguous processing of the entries and the management of the links between all the objects of the file.

Example: encoding of an English entry

`<entry n="399" id="en399">`	*beginning of the entry: specific identifier*
`<form type="main"><orth type="main">base terminals</orth></form>`	*term that constitutes the object of the entry*
`<gramgrp><pos>n</pos></gramgrp>`	*grammatical information*
`<sense n="1" id="en399s1">`	*start of the 'defining' section, each meaning is numbered*
`<def>The two extremes of a base line. Usually marked on a monument or metal stake buried beneath a visible surface mark.</def>`	*the actual definition*
`<xr type="seealso"><ref target="en4283s1">reference mark</ref></xr>`	*reference to another entry, here 'see also'*
`</sense>`	*end of the defining section*
`</entry>`	*end-of-entry tag*

This stage resulted in the creation of three distinct files that were identical to the base dictionaries, but added with SGML tags. All the information that was there at the start had been preserved.

Other operations aimed at harmonizing the three versions were also carried out in order to create a homogeneous end-product:

- The elimination of all types of entries that posed an obstacle to the searching of indexed fields: entries that were permutated (e.g. *chart: nautical*) or dismembered (e.g. *carte* f *de base*) were replaced by the natural form (*nautical chart*, *carte de base*).
- The systematic addition of a grammatical feature for each term in the English and Spanish versions.

- The suppression in all versions of terms that were capitalized to indicate that they formed the object of an entry. This functionality has been replaced by the possibility to exploit a network of conceptual relations.
- The systematic addition of a field mention: the indication of a field was extended to the other languages. In the absence of a specificied field, a general field was allocated by default.
- The selective correction of errors, spelling mistakes, ambiguities… after consultation with the dictionary editors if such was required.
- Typographical homogeneization (capitals, symbols, etc.).

4.2 From the TEI format to the MARTIF format

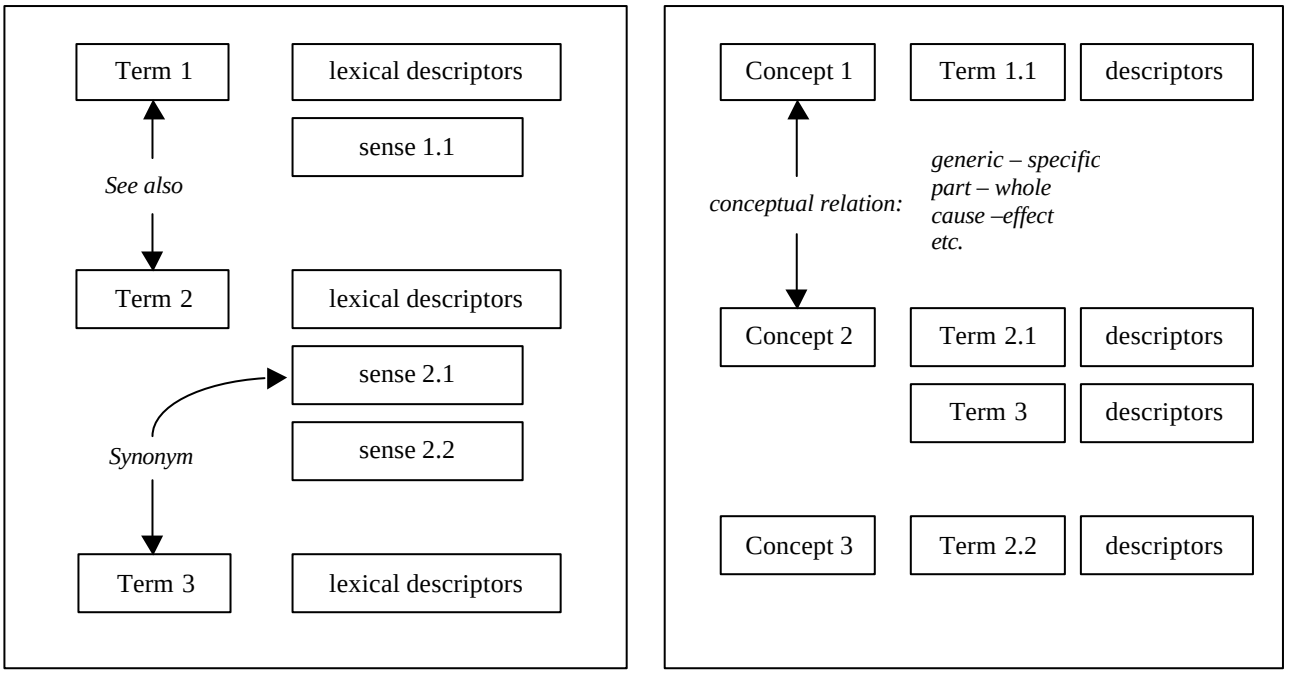
The aim was to develop a tool that would allow the conversion in one stage from the TEI format to the MARTIF format. It was necessary beforehand to 'prepare' the files. The next part contains a detailed discussion of the conversion method and any related tasks that were accomplished prior to starting the actual retroconversion.

4.2.1 FROM SPECIALIZED MONOLINGUAL LEXICOGRAPHY TO MULTILINGUAL TERMINOGRAPHY

For the transition from the initial polysemous model to the terminological monosemic model, it was convenient to group all lexical (term-related) and conceptual (related to each meaning of the term) information for each monolingual volume. Afterwards, it was necessary to combine the data belonging to one and the same concept and which were described in the various monolingual volumes, i.e. in the various languages.

The following diagram shows how these (lexical or semantic) 'information blocks' were transposed from one model to the other without the loss of any information. The various meanings of entries in the monolingual dictionaries to which synonyms classified alphabetically under other entries may refer, were separated in as many 'conceptual' entries, grouping together all the synonyms and their lexical descriptions. This shift from polysemy to monosemy necessarily implies the presence of homonyms in the new terminological data base.

References of the type 'see also' were from then on specified from concept to concept, independently of the languages, while constituting conceptual relations whose semantic scope may be specified: 'type of', 'part of', 'the opposite of', etc.



From the lexicological model to the terminological model
(monolingual view)

The successful transition from one model to another was founded on four hypotheses that were assumed valid, and on TEI-compliant encoding.

4.2.2 HYPOTHESIS 1: POLYSEMY OF THE DEFINITIONS

In each language, the various definitions attached to a single term do correspond to distinct meanings expressed by the term in question.

Example: distinct and autonomous definitions

```
<entry n="95" id="fr95">
<blocform>
<form type="main"><orth type="main">amarrage</orth></form>
</blocform>
<sense n="1" id="fr95s1">
<def>Action de mettre un navire à un poste fixe avec des amarres.</def>
</sense>
<sense n="2" id="fr95s2">
<def>Ensemble du dispositif d'amarres mises en place pour amarrer un navire.</ def>
</sense>
</entry>
```

As a result of the difficulty of verifying each individual definition some difficult cases of indistinct definitions (lack of a meaning) remained at this level of encoding. This is particularly true for the English version, in which the various meanings of the same entry are not separated by hard returns.

Example: three undifferentiated definitions for the same term

<entry n="85" id="en85">
<blocform>
<form type="main"><orth type="main">**alignment**</orth></form>
</blocform>
<sense n="1" id="en85s1">
<def>The placing of objects along a straight line. In navigation, the bringing into line of two or more conspicuous objects, such as lights, beacons, etc. Also their bearing as seen by an observer from seaward.</def>
</sense>
</entry>

Another scenario involves the existence of non-autonomous definitions, i.e. whose formulation depends on the form of other definitions. In the following example, the composition of definition No. 2 depends on definition No. 1. In the shift towards the conceptual model, definition No. 2 attached to the French term *dessin* is not relevant and is thus subject to revision.

<entry n="1327" id="fr1327">
<blocform>
<form type="main"><orth type="main">**dessin**</orth></form>
</blocform>
<sense n="1" id="fr1327s1">
<def>**Procédé consistant à réaliser une image** par dépôt sur un support d'une substance opaque, à l'aide d'outils appropriés.</def>
</sense>
<sense n="2" id="fr1327s2">
<def>**Image réalisée par le procédé précédent**, ou l'ensemble du support et de l'image.</def>
</sense>
</entry>

These residual cases will have to be located and corrected by the editors themselves with the aid of the editing tool that has been developed for them.

4.2.3 HYPOTHESIS 2: EQUIVALENCE OF MEANINGS BETWEEN THE LANGUAGES

It is essential that the numerical references that point the French entries to the English entries should be directly transposable to meaning-to-meaning equivalence relations.

In 2.2 we have seen that the meanings are not numbered in the English volume and that each meaning of the French volume refers to an English term and not to a specific meaning of the term in question. To resolve this, the help of the French language editors was enlisted in the course of what was a highly difficult phase of disambiguating the references from the polysemous entries. For each case that was identified, their task consisted of indicating which of the English meaning(s) was (were) actually related to the initial reference. Thanks to this painstaking task, the current references are now all of the 'meaning-to-meaning' type, and thus fulfil the criteria of hypothesis No. 2.

4.2.4 HYPOTHESIS 3: PRINCIPLE OF CONCEPTUAL EQUIVALENCE

If A of L_1 (language 1) is equivalent to a of L_2 (language 2) and if A of L_1 is equivalent to ß of L_2, whereas a of L_2 is not a synonym of ß of L_2, then A of L_1 probably has two meanings that should be differentiated by means of two distinct entries within the dictionary. (Van Campenhoudt 1996)

So, in accordance with the above hypothesis, a homonymous entry has been created for the English term *drizzle*, which may be translated into French as *bruine* or *crachin*, which are not actually synonyms in French [example 3].

Bruine – *n.m.*	**drizzle** – *n*
Définition:	Definition:
Précipitation de densité assez uniforme, constituée exclusivement de fines gouttelettes d'eau (souvent froide, de diamètre inférieur à 0,5 mm), très rapprochées les unes des autres.	Fairly uniform precipitation composed exclusively of fine drops of water (diameter less than 0.5 mm), very close to one another.
Voir aussi: crachin	See also: drizzle
crachin - *n.m.*	**drizzle** – *n*
Définition:	Definition:
Pluie très fine de caractère spécifiquement marin. N'implique pas l'idée de froid que comporte la bruine.	Fairly uniform precipitation composed exclusively of fine drops of water (diameter less than 0.5 mm), very close to one another.

A list of these cases was drawn up in order to enable the French and English language editors to check the relevance of the results.

4.2.5 HYPOTHESIS 4: SEMANTIC LINKS BETWEEN MONOSEMIC MEANINGS

The internal references in each language that indicate a semantic relation between entries (and thus implicity between a set of terms) are immediately transposable into relations of the same type between meanings.

In terms of its complexity, this problem is rather similar to the one raised in hypothesis No. 2. The French version almost systematically provides internal meaning-to-meaning references, whereas the English version almost exclusively has internal meaning-to-entry references, which can once again be explained by the the fact that meanings are not numbered.

For each of the languages, the internal references (pointing towards a polysemous entry) that could not be resolved automatically were submitted to the editors involved so that from now on they can be interpreted as real semantic relations.

4.3 Technical realization: a reformatting tool

The same IT tool was used in the course of the various stages in the transformation of the data model and the encoding format. Written in Java by LORIA, the program exploits the entry data of the three SGML files corresponding to the French, Spanish, and English monolingual dictionaries encoded in accordance with TEI recommendations. This application does not require any manual intervention and generates a set of XML files in MARTIF format, with each corresponding to a particular concept.

The processing was done in five consecutive phases.

Phase 1: examination of all meanings, starting from the French dictionary. This phase allowed the identification and creation of four types of concepts:

- concepts described in the three languages: French, English, and Spanish;
- concepts described in French and in English;
- concepts described in French and Spanish;
- concepts described only in French.

Phase 2: examination of all meanings contained in the English version that had not been taken into account during phase 1. This resulted in the creation of two types of concepts:

- concepts described in English and in Spanish;
- concepts described only in English.

Phase 3: examination of all meanings in the Spanish version that had not been taken into phases 1 and 2. This phase allowed the identification of concepts described only in Spanish.

Phase 4: creation of all the conceptual links (exploitation of hypothesis 4, cf. 4.2.5) by extending the semantic links thitherto encoded within one language to the concept level. The resulting mechanical solution will undoubtedly require careful checking by the language editors themselves.

Phase 5: renumbering of the concepts, and, consequently, of the semantic links created during phase 4.

In the course of these various processes, the application generates a set of lists to be exploited systematically by the editors: lists of meanings mentioned in only one language, or two languages out of three, list of triplets (number of the French meaning, number of the English meaning, identifying the concept created), etc.

The data base created at the end of the above process counts 5,349 concepts, divided as follows:

	trilingual	**bilingual**			**monolingual**		
	Fr-En-Es	**Fr-En**	**Fr-Es**	**En-Es**	**Fr**	**En**	**Es**
Number	*4 684*	*5*	*0*	*261*	*362*	*27*	*10*
Percentage	*87.57 %*	*0.09 %*	*0.0 %*	*4.88 %*	*6.77 %*	*0.50 %*	*0.19 %*

A close examination of this table results in the following observations and hypotheses:

- The majority (over 87 %) of the concepts identified are described in the three languages of the *Hydrographic Dictionary*, which confirms the relevance of the conceptual model that was chosen.
- The fact that no bilingual French-Spanish concept has been identified confirms that the Spanish version has been entirely modelled on the English version (cf. note 5).
- The high number of concepts described only in French (362) would seem to indicate that the French editorial group makes finer distinctions between meanings. Yet, the equally large quantity of bilingual English-Spanish concepts (261) seems to favour the hypothesis that a number of links remains to be expressed between French and English. The lists of those bilingual concepts will be of great assistance to the editors in their task of validating one or other of the two hypotheses.
- Once the corrections have been carried out, it is likely that there will still be a small number of concepts that are described monolingually and which express terminological specificities peculiar to each language.

5 Exploitation of the results

5.1 XSL and the stylesheets

Using the XML file resulting from the retroconversion as a basis, it is possible to extract all kinds of products thanks to the XSLT (*XSL Transformations*) mechanism which enables the transformation of one XML document into another XML document by means of a stylesheet containing all the layout transformation rules.

As a result, XSL files, combined with the filter mechanisms, will allow the customized extraction and sorting of lists of objects (terms, concepts, synonyms, etc.) in accordance with various criteria (field, date, language, etc.). Examples include:

- Alphabetically sorted multilingual dictionaries based on a source language;
- Polysemous monolingual dictionaries similar to the old printed versions;
- Monolingual or bilingual glossaries containing only the terms and their equivalents (without definitions);
- The list of all concepts modified at a certain date;
- Lists of results of well-defined requests when consulting the dictionary.

5.2 Prospects

The structuring and encoding of the information contained in the *Hydrographic Dictionary*, its continuous updating by the editors only make sense if it is made available to users. Hence, the following stage will consist of exploiting the contents along two axes:

- The possibility by the IHO to distribute a variety of derived products thanks to the XSL modules. For instance, a bilingual thematic glossary aimed at conference interpreters.
- An Internet consultation interface for replying to user requests.

6 Conclusion

In accordance with the hypotheses by Marc Van Campenhoudt (2000: 139-140), which are included here, the results of the *Dhydro* project reveal that thanks to a careful identification of the contents of the dictionaries it is possible to transform monolingual polysemy-based lexicographical works into a monosemy-based multilingual data base.

Moreover, this would lead one to think that within a computer context the demand for an accurate identification of the contents and for the abolition of the linear writing-based arrangement will result in a pragmatic reconsideration of the already blurred borders of specialized lexicography and terminography. In front of the computer, the lexicographer is on an equal footing with the terminographer who has to distribute information within predefined fields.

There is nevertheless an intractable obstacle between immediately applicable theory and practice; a great many authors of specialized dictionaries continue to design their work in accordance with the editorial models of the great dictionaries of the past. Modern computing technology will result in a real revolution in the approach to dictionary making: from now on, dictionary authors must think in terms of data bases rather than in terms of typography and page layout. They have to learn to work out strict data models; the more information – particularly semantic information – is defined with precision and accuracy, the more varied the use of the dictionary will be in a society that is already multilingual.

What is more, the quality of tagging tools and with it new prospects for direct retrieval are elements in favour of the creation of specialized ‘metadictionaries’ which are capable of generating a large variety of derived products for one or several languages.

References

BESSERO (G.), BONHOMME (P.), ROMARY (L.), 1999: "L'informatisation du dictionnaire hydrographique international", dans Newman (D.L.) et Van Campenhoudt (M.), eds, *Maritime Terminology: Issues in Communication and Translation. Proceedings of the First International Conference on Maritime Terminology (Brussels, 15-16 May 1998)*, Brussels, Editions du Hazard, p.178-198.

CENTRE DE RECHERCHE TERMISTI & LORIA, 1999: *Le modèle de données et la D.T.D.*, D 1.2 of the MLIS-2009 *Dhydro* project, March 1999.

IDE (N.) et VÉRONIS (J.), 1996: "Codage TEI des dictionnaires électroniques", *Cahier GUTenberg, Numéro spécial: TEI - Text Encoding Initiative*, June 1996, No. 24, p. 170-176.

INSTITUT FÜR DEUTSCHE SPRACHE, 1999: *Text Preparation and Alignment Work for DHYDRO*, deliverable D 3.1 du projet MLIS-2009 *Dhydro*, November 1999.

INTERNATIONAL HYDROGRAPHIC ORGANIZATION (1994): *Hydrographic Dictionary*, 5th ed., Monaco, Bureau hydrographique international.

ISO 12200 (1999): *Computer applications in terminology - Machine-readable terminology interchange format (Martif) - Negotiated interchange*, Geneva, International Organization for Standardization.

ISO 12620 (1999): *Computer applications in terminology - Data categories*, Geneva, International Organization for Standardization.

LORIA & CENTRE DE RECHERCHE TERMISTI, 1999: *Le modèle éditorial du forum Dhydro*, deliverable D1.1 of the MLIS-2009 *Dhydro* project, March 1999.

LORIA & CENTRE DE RECHERCHE TERMISTI, 1999: *Choix techniques d'accès aux données, échanges et outils*, deliverable D2.1 of the MLIS-2009 *Dhydro* project, March 1999.

LORIA & CENTRE DE RECHERCHE TERMISTI, 1999: *Le D.H.I sous format S.G.M.L.*, deliverable D 1.3 of the MLIS-2009 *Dhydro* project, December 1999.

ORGANISATION HYDROGRAPHIQUE INTERNATIONALE (1998): *Dictionnaire hydrographique*, 5th ed., Monaco, Bureau hydrographique international.

ORGANIZACIÓN HIDROGRÁFICA INTERNACIONAL (1996): *Diccionario Hidrográfico*, 5th ed., Monaco, Bureau hydrographique international.

VAN CAMPENHOUDT, M. (1996): "Réseau notionnel, intelligence artificielle et équivalence en terminologie multilingue: essai de modélisation", dans Clas, A., Thoiron, Ph. et Béjoint, H., éd., *Lexicomatique et dictionnairiques, IV^es journées scientifiques du réseau thématique « Lexicologie, terminologie, traduction », Université Lumière (Lyon II), 28-30 septembre 1995*, Montreal, AUPELF-UREF and Beirut, F.M.A., p. 281-306.

VAN CAMPENHOUDT, M. (2000): "De la lexicographie spécialisée à la terminographie: vers un « métadictionnaire »?", in Thoiron (Ph.) et Béjoint (H.), dir., *Le sens en terminologie*, Lyon, Presses universitaires de Lyon (Travaux du C.R.T.T.), p. 127-152.

Annex: examples from the English and French volumes

[example 1]

Concordance. *f*
a) Traitement des observations de MARÉES consistant à rechercher les paramètres de corrélation entre les heures et hauteurs de PLEINE et BASSE mers d'un lieu et les éléments correspondants d'un autre lieu.
b) Résultat de l'opération précédente.

[example 2]

987 **Constellation**. Formerly a number of FIXED STARS grouped within an imaginary outline. Now, a region of the CELESTIAL SPHERE marked by arbitrary boundary lines.

Constellation. *f*
987 a) À l'origine, groupe d'ÉTOILES FIXES, comprises à l'intérieur du contour d'une figure imaginaire. Actuellement, région de la SPHÉRE CÉLESTE arbitrairement délimitée par des arcs de MÉRIDIENS et PARALLÈLES célestes.
b) Groupe de SATELLITES ARTIFICIELS appartenant à un même système, tel que GPS.

[example 3]

1479 **drizzle.** Fairly uniform PRECIPITATION composed exclusively of fine drops of water (diameter less than 0.5 mm), very close to one another.
1479 **Bruine.** *f*
PRÉCIPITATION de densité assez uniforme, constituée exclusivement de fines gouttelettes d'eau (souvent froide, de diamètre inférieur à 0,5 mm), très rapprochées les unes des autres. Voir aussi CRACHIN.
1479 **Crachin.** ***m***
Pluie très fine de caractère spécifiquement marin. N'implique pas l'idée de froid que comporte la BRUINE.

[example 4]

1907 **foreshore.** That part of SHORE which lies between HIGH and LOW WATER MARK at ordinary TIDE.
5074 **strand.** The portion of the SEASHORE between high and low water line.

1907-5074 **Estran.** *m*
Partie du RIVAGE comprise entre la LAISSE DE HAUTE MER et la LAISSE DE BASSE MER.

[example 5]

214 **apron.**
1. A gently dipping featureless surface, underlain primarily by SEDIMENT, at the base of any steeper SLOPE.
2. The area of wharf or quay for handling cargo.
3. A sloping underwater extension of an iceberg.
4. An outwash plain along the front of a glacier.

Extract from the index of the French version, which allows the user to find all the French terms for which *apron* is given as an equivalent:

214 Aire de manutention
Éperon (sous-marin) c)

Glacis

214 **Aire *f* de manutention.**
Partie supérieure d'un APPONTEMENT ou d'un QUAI destinée à la manutention des marchandises.

Éperon *m* (sous-marin).
a) Partie saillante sous-marine d'un CAP.
4942 b) Partie en saillie d'un ensemble morphologique plus important, PENTE, DORSALE, MASSIF etc.
214-4207 c) Partie saillante immergée d'un MUR DE GLACE, d'une FALAISE DE GLACE, d'un ICEBERG ou d'un FLOE.

214 **Glacis**. *m*
Surface unie, de faible PENTE a) et de genèse essentiellement sédimentaire à la base d'une PENTE b) plus forte.

[example 6]

4726 **signal: eccentric.** See ECCENTRIC SIGNAL.
4727 **signal: fog.** See FOG SIGNAL.
4728 **signal: hydrographic.** Any object, existing or specially erected, for the CONTROL of HYDROGRAPHIC SURVEY. See also SOUNDING MARK.
4729 **signal: longitude.** A sign indicating a time event, observable at different stations, and used in comparing LOCAL TIMES of those stations, and determining the difference of their LONGITUDES.
4730 **signal: luminous.** In SURVEYING, a SIGNAL consisting of HELIOTROPE or HELIOSTAT by day, and SELENOTROPE or SIGNAL LAMP by night.
4731 **signal: sound.** See SOUND SIGNAL.

[example 7]

5066 **storm surge.** A RISE above normal water level on the OPEN COAST due only to the action of WIND stress on the water surface. Storm surge resulting from a HURRICANE or other intense STORM also includes the RISE in level due to ATMOSPHERIC PRESSURE reduction as well as that due to wind stress. A storm surge is more severe when it occurs in conjunction with a HIGH TIDE. Also called *storm tide, storm wave, tidal wave.*
5067 **storm tide.** See STORM SURGE.
5918 **wave: storm.** A wind-generated sea surface wave of great HEIGHT. See STORM SURGE.

5066-5067 **Marée *f* de tempête.** Voir ONDE DE TEMPÊTE.
5918 **Onde *f* de tempête.**
Élévation anormale du NIVEAU DE LA MER, d'une durée de quelques heures à quelques jours, provoquée par une action violente du VENT sur la surface de la mer et/ou par la baisse de la PRESSION ATMOSPHÉRIQUE. Elle se propage comme une ONDE DE GRAVITÉ et peut se renforcer dans des EAUX PEU PROFONDES. On dit aussi *marée de tempête*.

[example 8]

101 **altitude.** The vertical distance of a LEVEL, a point or an object considered as a point, measured from a given datum, usually MEAN SEA LEVEL. See also ELEVATION. In ASTRONOMY, the vertical angle between the plane of the HORIZON and the line to a CELESTIAL BODY. See also ANGLE OF DEPRESSION and ANGLE OF ELEVATION.
1590 **elevation.** The vertical distance of a point or a LEVEL, on or affixed to the surface of the EARTH, measured from MEAN SEA LEVEL. The term elevation is sometimes used synonymously with ALTITUDE which in modern use refers particularly to the distance of points or objects above the EARTH's surface.
An area higher than its surroundings, as a HILL.
2223 **height.** The vertical distance of a LEVEL, a point, or an object considered as a point, measured from a

specified DATUM.
The vertical dimension of an object.

Hauteur. *f*

2223 a) Dimension verticale d'un objet.

101-2223 b) Distance verticale entre un point et un NIVEAU DE RÉFÉRENCE. On dit aussi ALTITUDE. Dans le cas d'un FEU la HAUTEUR désigne l'altitude du foyer par rapport au sol, l'altitude ou ÉLÉVATION désigne l'altitude du foyer par rapport à un niveau de pleines mers.

1590-5764 c) Zone plus élevée que son voisinage, telle une COLLINE.

101 d) En ASTRONOMIE, angle de la direction d'un ASTRE avec le plan horizontal. Voir aussi ANGLE DE DÉPRESSION et ANGLE D'ÉLÉVATION.

Élévation. *f*

1590 a) Partie d'un TERRAIN dont l'ALTITUDE est supérieure à celle du terrain environnant, par exemple une colline.

2223 b) Synonyme de HAUTEUR a) ou b).

c) Dans le cas d'un FEU, altitude de la lanterne au-dessus du niveau des PLEINES MERS MOYENNES DE VIVE-EAU (usage français).

[example 9]

5522 **time: Greenwich mean (GMT) (or Greenwich civil).** MEAN SOLAR TIME at the GREENWICH MERIDIAN. See UNIVERSAL TIME (UT).

6064 **Zulu time.** See TIME: GREENWICH MEAN.

Heure. *f*

2309 **a)** Unité de temps égale à la vingt-quatrième partie du JOUR c). Le symbole est h.

5513 **b)** Moment déterminé du jour a) rapporté à une origine temporelle bien déterminée. Dans ce sens le mot heure entre dans la désignation des échelles de temps (voir termes ci-après).

c) Heure suivie d'une lettre capitale désigne l'heure d'un FUSEAU HORAIRE. Ainsi HEURE Z (*Zoulou*) désigne le TEMPS UNIVERSEL.

6064 **Heure *f* Zoulou.** Voir HEURE c).
